Semantic Search and Discovery

CENDI, January 12, 2012 @ LC

Tamas Doszkocs, Ph.D.

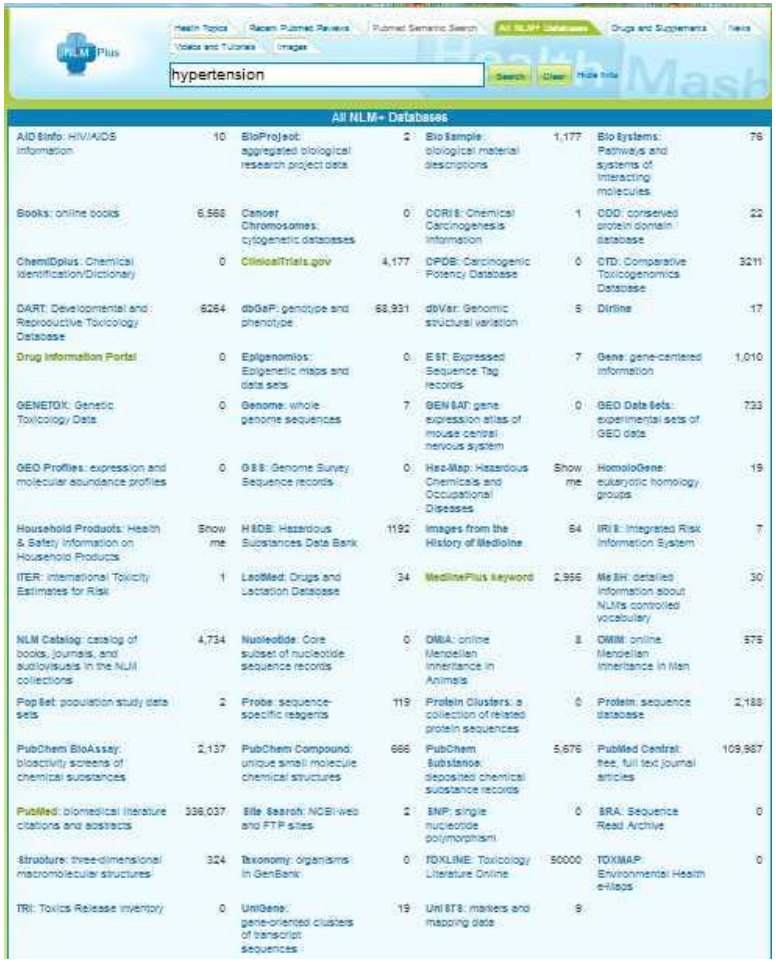
tamas@weblib.com

What is semantic search?

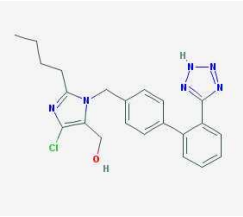
Keywords vs. Concepts

"**semantic search** is **a search** or **a question** or **an action** that produces **meaningful results**, even when the retrieved items contain **none of the query terms**, or the search involves **no query text** at all "

What is NLMplus?



NLMplus is a **semantic search** and **knowledge discovery** application designed to tap into the **rich content offerings** of the **National Library of Medicine** in all areas of **biomedicine** and **health**



adhd aids alt amlodipine anemia anxiety arthritis aspirin asthma atenolol atrial fibrillation autism bipolar blood pressure breast cancer bun calcium cancer cbc celexa cellulitis cholesterol citalopram clonazepam clonidine colonoscopy copd coumadin creatinine cystic fibrosis dementia depression diabetes digoxin diovan diverticulitis diverticulosis eczema endometriosis fibromyalgia furosemide gabapentin gerd glaucoma gout heart hematocrit hemoglobin hepatitis hepatitis c hernia hiatal hernia hiv hydrochlorothiazide hyperlipidemia hypertension hypothyroidism interactive clinical trial tutorial iron leukemia lipitor lisinopril liver lorazepam lung cancer lupus magnesium melatonin metformin metoprolol mrsa multiple sclerosis niacin obesity omeprazole osteoarthritis osteoporosis pancreatitis pneumonia potassium prednisone pregnancy pyelonephritis rheumatoid arthritis risperidone sarcoidosis scabies schizophrenia scoliosis seroquel shingles simvastatin stroke synthroid thyroid tramadol vitamin c vitamin d vitamin e zinc

NLMplus Background

- NLMplus is an award winning application of **semantic search** and **knowledge discovery** technology which was developed by WebbLib LLC during the first year of a DOE/OSTI SBIR project

- NLMplus addresses the difficult problem of **accessing heterogeneous legacy databases** with a meaningful solution in order to produce **improved relevance and precision in search results**

- The **goal** of the SBIR project was to research and implement innovative semantic tools for the improved indexing, searching and dissemination of **scientific and technical information.**

- As part of the SBIR research, WebLib had implemented a Green Energy Semantic Search pilot, which was subsequently deployed into production use by OSTI

- Where possible, WebLib's semantic solution **enriches** and utilizes **existing thesauri** and **ontologies** in a Web Knowledge Base which powers the semantic search.

  - Typical thesaurus problem: "**green energy**" is **NOT** in the ETDE/INIS Joint Energy Thesaurus

Semantic Annotation of PubMed Reviews

anthracycline cardiotoxicity after breast cancer treatment

Display Settings: Abstract

Send to:

Oncology (Williston Park). 2009 Mar;23(3):227-34.

Anthracycline cardiotoxicity after breast cancer treatment.

Hershman DL, Shao T.

Columbia University, New York, N Y 10032, USA.

Abstract

Anthracyclines are among the most active agents for the treatment of breast cancer; their use in combination regimens improves both disease-free and overall survival in patients with breast cancer. Unfortunately, the clinical utility of anthracycline use is limited by a cumulative dose-dependent cardiac toxicity resulting in congestive heart failure. As methods for detecting and treating breast cancer improve, there has been a steady decline in breast cancer mortality over the past 15 years. With an increasing number of long-term breast cancer survivors, the number of patients experiencing anthracycline-induced cardiotoxicity may also continue to grow. Moreover, new agents used in the treatment of breast cancer can potentiate cardiac toxicity. Recently, studies of non-anthracycline-containing regimens have been found to be effective in preventing recurrence of breast cancer (as compared with anthracycline-containing regimens) in patients with early-stage breast cancer, with a reduced incidence of adverse cardiac outcomes. In this article, we summarize the incidence, presentation, and mechanism of anthracycline-associated cardiotoxicity. We also discuss risk factors for the development of anthracycline-induced cardiotoxicity and new therapies, such as trastuzumab, that may potentiate cardiac toxicity. Finally, we review monitoring and preventive practices that may reduce the long-term risk of anthracycline-related cardiotoxicity.

Semantic Annotation of a PubMed Review Title

via the WebLib Part-of-Speech Tagger and Noun Phrase Chunker

phrases> **key phrases**> key concepts > synonyms > normalized semantic indexes

19418823;TI:anthracycline cardiotoxicity after breast cancer treatment

TI_NORMALP: anthracycline breast canc cardiotoxicit treat

TP:**anthracycline**;

- CUI:C0003234; Anthracycline Antibiotics;
- SY: anthracyclines;

– TP:**cardiotoxicity**;

– CUI:C0876994; cardiac toxicity;

– TP:**breast cancer**;

- CUI:C0678222; Breast Carcinoma;
- SY: breast tumor;

– TP:**breast cancer treatment**;

- CUI:C1511300; Breast Cancer Therapeutic Procedure;

Scaling the Semantic Searching of PubMed Reviews

- More than **1.6 million PubMed Reviews** have been **semantically indexed** on a small WebLib server
  - Query: **prostate cancer antigen 3**
  - Query: **transient cortical blindness after coronary angiography**

Health Topics | **Pubmed Reviews Semantic Search** | Enhanced Pubmed Search | All N

Drugs and Supplements | News | Videos and Tutorials | Images

prostate cancer antigen 3 **Search** **Clear**

Related Concepts: dd3, prostate-specific gene dd3, pca3, pca3 gene, prostate cancer a coding rna 19

Latest Reviews from Pubmed 0 *NLMplus Pubmed Reviews Semantic Search* 40 (from 40)

40 top results from 40 found for: prostate cancer antigen 3...

Pubmed Reviews Semantic Search Results

1 2 3 4 Next >

[Assessment of the PCA3 test for prostate cancer diagnosis: a systematic ...

"upm3", "**PCA3**", "dd3", "aptima **PCA3**" and "**prostate cancer antigen 3**". Patients were adults. The intervention was to determine the **PCA3 gene**, from urine samples for diagnosis of **prostate cancer**. The quality ... OBJECTIVE: To assess the efficacy of diagnostic techniques based on **PCA3 gene** for early detection
[Ruiz-Aragon J, Marquez-Pelaez S] - [Actas Urol Esp, 2010; 34(4):346-355]

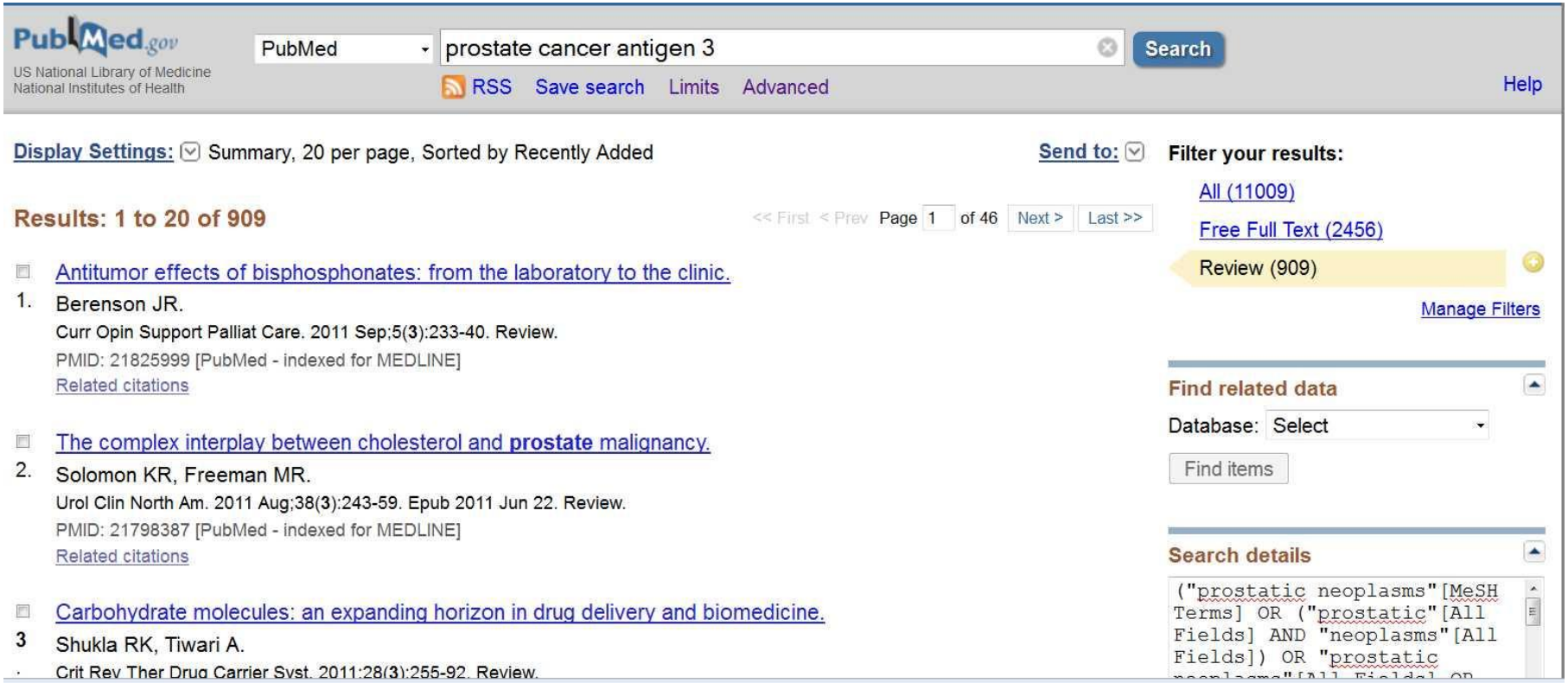
PCA3: from basic molecular science to the clinical lab.

are emerging and **prostate cancer gene 3** (**PCA3**) is one such marker. **PCA3** is a noncoding RNA that is highly over-expressed in **prostate cancer** tissue compared to benign tissue. A non-invasive test for **PCA3** was developed ... demonstrated the utility of **PCA3** for the diagnosis of **prostate cancer** and some studies suggest that **PCA3** may
[Day JR, Jost M, Reynolds MA, Groskopf J, Rittenhouse H] - [Cancer Lett., 2011; 301(1):1-6]

Progensa PCA3 test for prostate cancer detection.

are required in order to avoid unnecessary biopsies. The **PCA3 gene** product is specifically overexpressed ... in order to

Search Results from NLM's PubMed Search Engine



4 top results from 4 found for: transient cortical blindness after coronary angiography...

Pubmed Reviews Semantic Search Results

Transient cortical blindness after coronary angiography: a ...

Transient cortical blindness after coronary angiography is a rarely encountered, rapidly ... study. We report a case of **transient cortical blindness** following diagnostic **coronary angiography** ... include those undergoing **coronary** artery bypass graft study and those with renal failure. Although
[Akhtar N, Khatri IA, Naseer A, Ikram J, Ahmed W] - [J Pak Med Assoc, 2011; 61(3):295-297]

Cortical blindness: a rare but dramatic complication following coronary angiography.

[Gellen B, Remp T, Mayer T, Milz P, Franz WM] - [Cardiology, ; 99(1):57-59]

An uncommon complication after a common procedure.

A 50-year-old female developed **cortical blindness** 5 hours **after coronary angiography**. A non-contrast-enhanced computed tomographic scan of the brain revealed bilateral occipital and middle cerebral/posterior cerebral artery watershed territory infarcts with a possibility of subarachnoid hemorrhage (SAH). Magnetic
[Math RS, Singh S, Bahl V] - [J Invasive Cardiol, 2008; 20(10):E301-E303]

Transient cortical blindness following bypass graft angiography. A case report.

Transient cortical blindness, an uncommonly recognized complication of cerebral **angiography** ... -old patient who had **transient cortical blindness** following bypass graft **angiography**. In this case ... the volume injected. When **cortical blindness** occurs, it is a serious clinical problem whether **transient**
[Kamata J, Fukami K, Yoshida H, Mizunuma Y, Moriai N, Takino T, Hosokawa S, Hashimoto K, Nakai K, Kawazoe K, et al] - [Angiology, 1995; 46(10):937-946]

PubMed Query: **transient cortical blindness following coronary angiography**

Display Settings: Summary, Sorted by Recently Added

Send to:

Results: 2

1. **Transient cortical blindness** after **coronary angiography**: a case report and literature review.
   Akhtar N, Khatri IA, Naseer A, Ikram J, Ahmed W.
   J Pak Med Assoc. 2011 Mar;61(3):295-7. Review.
   PMID: 21465952 [PubMed - indexed for MEDLINE]
   Related citations

2. **Transient cortical blindness following** bypass graft **angiography**. A case report.
   Kamata J, Fukami K, Yoshida H, Mizunuma Y, Moriai N, Takino T, Hosokawa S, Hashimoto K, Nakai K, Kawazoe K, al.
   Angiology. 1995 Oct;46(10):937-46. Review.
   PMID: 7486215 [PubMed - indexed for MEDLINE]
   Related citations

The WebLib Semantic Search Engine

- **Apache Solr open source enterprise search engine with WebLib's Semantic Layer**
  - Please note: WebLib's semantic indexing and search technology can be implemented for any enterprise search engine, e.g. the **MarkLogic** system, which is being increasingly utilized by federal agencies

- WebLib's **Biomedical Knowledge Base** and **Query Translator**

- WebLib's **Web Knowledge Base** Prototype

More about the NLMplus

NLMplus Innovations

- Universal Search
  - Flexible access to 60 NLM databases
- Semantic Search Engine
  - Semantic indexing and searching of PubMed Reviews
- Biomedical Knowledge Base
  - powers “Explore and Discover”
  - powers the Semantic Search Engine
- Intuitive User Interface
- Wide Spectrum of NLM users
  - Consumers
  - Physicians
  - Biomedical researchers
  - Decision makers

NLM APIs, web services and software tools utilized by the NLMplus app

- Entrez Programming Utilities
- UMLS Terminology Services
- Medlineplus Web Service
- MetaMap API
- SKR Web API

Potential WebLib Semantic Search and **Knowledge Base Applications**

Common Problem: ***accessing heterogeneous databases***

Problems in Keyword Searching: ***relevance and precision***

Meaningful Solution: ***semantic technologies***

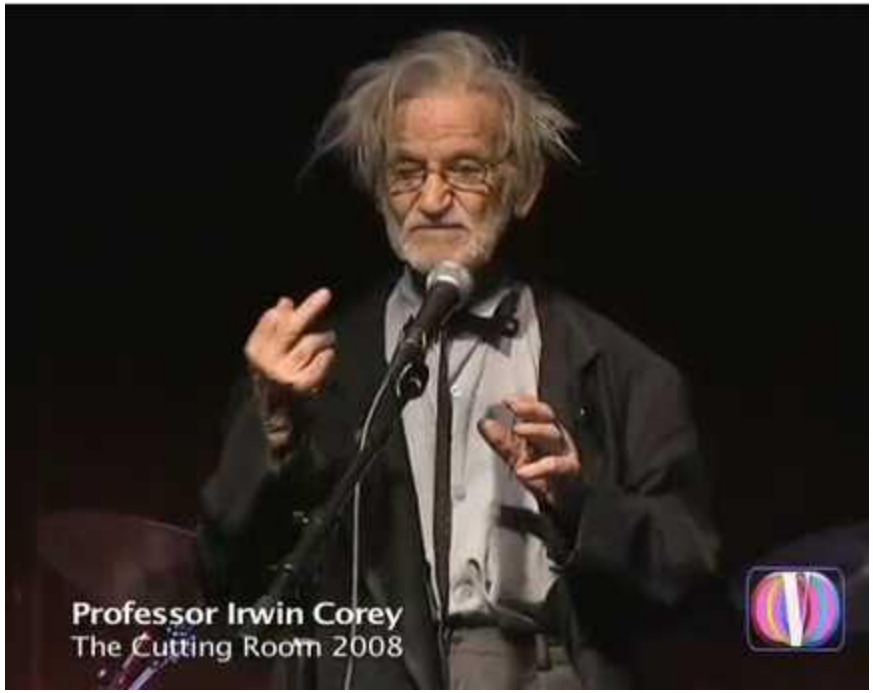
Biomedical Knowledge Base
Semantic Search

- National Library of Medicine
- National Institutes of Health
- Department of Health and Human Services
- Medical Schools
- Health Science Libraries
- Health Care Industry

Web Knowledge Base
Semantic Search

- U.S. Government Departments and Agencies
- Libraries and Research Organizations
- Media and Content Companies
- Educational Institutions
- Businesses
- The Web

Professor Irwin Corey Explains Semantics



Semantic Search and Discovery

CENDI, January 12, 2012 @ LC

Tamas Doszkocs, Ph.D.

tamas@weblib.com